STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **16**, October 2026, pp 3669–3682.
Published online in International Academic Press (www.IAPress.org)

Machine Learning-Based Volatility Forecasting and Systemic Risk Dynamics in Indonesian State-Owned Banks

Nono Heryana[1], Nugraha Nugraha[2], Maya Sari[2,*], Disman Disman[2], Toni Heryana[2], Rini Mayasari[1]

[1] *Universitas Singaperbangsa Karawang, Karawang, Indonesia*
[2] *Universitas Pendidikan Indonesia, Bandung, Indonesia*

Abstract This study evaluates volatility forecasts and systemic-risk indicators for four Indonesian state-owned banks (BBRI, BBTN, BMRI, and BBNI) from January 2010 to December 2025. Random Forest (RF) and Gradient Boosting (GB) models use information available at each forecast origin and are tuned by expanding-window validation. We compare them on a common 774-day test sample with a Random Walk, historical mean, moving average, HAR, GARCH(1,1), EGARCH, and GJR-GARCH. The GARCH-family models produce the most accurate 20-day rolling-volatility forecasts, with out-of-sample R^2 values of 0.9793–0.9838. RF attains R^2 values of 0.8929–0.9459 and does not outperform persistence. HAC-corrected Diebold–Mariano tests favor the Random Walk over RF and GB, but favor the GARCH-family models over the Random Walk. The first principal component of the Systemic Risk Index (SRI) explains 76.7% of component variation and correlates 0.999 with the equal-weight index; ΔCoVaR estimates indicate downside dependence. After 10-basis-point switching costs, an ex-ante RF volatility-timing rule improves Sharpe ratios for BBRI, BMRI, and BBNI, but not BBTN. RF and GB reveal bank-specific nonlinear associations, whereas the econometric models provide the most accurate point forecasts.

Keywords Volatility forecasting; Systemic risk; Machine learning; Random Forest; Gradient Boosting; Rolling volatility; Interconnectedness; Indonesian banking; State-owned banks; Financial stability

AMS 2010 subject classifications 62P20, 91G70, 68T05

DOI: 10.19139/soic-2310-5070-3546

1. Introduction

Banks account for a substantial share of credit intermediation in emerging economies, making their risk exposures relevant to monetary transmission and financial stability [1, 2]. Indonesian state-owned banks also support government-linked financing and financial inclusion; market-risk shocks affecting these institutions may therefore have implications beyond individual shareholders [3]. Bank-level volatility forecasts can inform risk management and prudential monitoring, provided that their accuracy is assessed against credible benchmarks.

GARCH-type and stochastic-volatility models remain standard forecasting tools because they impose a coherent structure on conditional variance. Their parametric form may, however, be restrictive when volatility responds nonlinearly or changes across regimes [4, 16]. In banking systems, correlated exposures and institutional linkages can further connect bank-specific volatility to system-wide risk [5, 6].

Tree-based methods offer a less restrictive way to model nonlinearities and interactions in financial data [7, 8]. Evidence of their forecasting performance is mixed and depends on the benchmark, predictor design, and treatment of temporal ordering [9, 20]. Random data splits, contemporaneous predictors, and extensive tuning can overstate out-of-sample performance, while variable-importance measures do not by themselves establish an economically interpretable mechanism.

*Correspondence to: Maya Sari, Universitas Pendidikan Indonesia, Bandung, Indonesia (Email: mayasari@upi.edu).

ISSN 2310-5070 (online) ISSN 2311-004X (print)

Systemic-risk research uses correlation, spillover, and network measures to characterize dependence among financial institutions and identify potential shock transmitters [10, 11]. Volatility and co-volatility are central to these measures because simultaneous increases across institutions can accompany broader financial stress [12, 13]. This monitoring role is distinct from prediction: an indicator may describe contemporaneous stress without improving a one-day-ahead bank-volatility forecast.

The empirical question is whether nonlinear models and system-wide variables add information beyond the latest volatility observation, and whether any predictive gain has economic value. We address it using strictly lagged predictors, time-ordered validation, a composite SRI interpreted as a contemporaneous monitoring measure, and bank-specific volatility timing. The timing exercise follows prior work that evaluates volatility forecasts through portfolio performance [35]; temporal ordering and training-only parameter selection prevent look-ahead information from entering the evaluation. Inference is limited to the four listed state-owned banks: persistence remains difficult to beat, systemic indicators are more informative for monitoring than for one-day-ahead prediction, and economic performance differs across institutions.

2. Related Work

2.1. *Volatility Forecasting in Financial Markets*

GARCH-type and stochastic-volatility specifications remain standard benchmarks in risk measurement and asset allocation. Their extensions accommodate asymmetry and exogenous predictors, but model performance can deteriorate when the variance process changes across regimes [14, 15]. Comparative evidence from developed and emerging markets consequently treats persistence and structural change as central forecasting challenges [16, 17].

Machine-learning models relax some parametric restrictions and can represent interactions among predictors. Their success in empirical asset pricing does not ensure superior volatility forecasts, because performance depends on the information set and evaluation design [18, 19]. Christensen et al. show that machine learning can forecast volatility effectively under a temporally ordered design [20]. Applications to Egyptian equities and XGBoost time-series forecasting further demonstrate that reported performance is specific to the market, predictor set, and validation procedure [21, 22].

2.2. *Systemic Risk and Interconnectedness in Banking*

Systemic risk cannot be inferred solely from the marginal risk of individual institutions. Connectedness measures instead represent how shocks may propagate through correlations, variance spillovers, or network links [10, 11]. Common movements in financial vulnerability provide related evidence that institution-level conditions can coincide with aggregate stress [23].

Market-based systemic-risk measures often incorporate common volatility because it captures changes in dependence that institution-specific risk measures omit [5, 6]. Concentrated banking networks and liquidity fragility in emerging markets can shape these dependencies [24, 25]. Yet the ability of a market-based stress measure to summarize contemporaneous dependence does not establish that it adds incremental forecasting information.

2.3. *Machine Learning and Financial Stability*

Machine learning has been applied to banking-crisis warnings and high-dimensional tail-risk assessment [26, 27]. Its capacity to approximate nonlinear relationships is useful when feedback effects are difficult to specify parametrically [28]. The same flexibility complicates interpretation and increases the risk of selecting unstable relationships, particularly when many predictors and model configurations are evaluated [29].

In banking applications, volatility-network measures summarize connectedness and may identify institutions whose movements precede changes elsewhere in the system [10, 11]. Machine-learning models can use such information to rank potential transmitters, although predictive importance does not establish causal contagion [30]. Evidence for narrowly defined state-owned banking segments in emerging markets remains limited.

2.4. Research Gap

Recent SOIC studies compare XGBoost and LSTM through explicit backtesting [31], combine realized-GARCH and HAR components with recurrent networks [32], and evaluate ARIMA, SVR, and LSTM across international indices [33]. The relevant gap is therefore not the absence of machine learning from financial forecasting. It is the lack of evidence on whether nonlinear models improve on persistence within a concentrated emerging-market state-owned banking segment, and whether contemporaneous stress variables add information to one-day-ahead bank-volatility forecasts.

Earlier forecasting studies primarily examine broad equity samples, international indices, or hybrid model design. The present analysis differs by focusing on a defined state-owned banking segment and evaluating point forecasts, contemporaneous systemic stress, and bank-specific economic outcomes within one empirical design. These distinctions are summarized in Table 1.

Table 1. Positioning Relative to Closely Related Studies

Study	Market/data	Main methods	Distinction from the present study
Christensen et al. [20]	29 U.S. equities	Multiple ML models and HAR benchmark	Developed-market forecast comparison without a state-owned-bank systemic-risk application
Oukhouya et al. [31]	Six international stock indices	XGBoost, LSTM, hybrid model, backtesting	Forecasts prices and trends across indices rather than bank rolling volatility and systemic stress
Bakkali et al. [32]	Persistent and anti-persistent financial series	RGARCH/HAR and recurrent-network hybrids	Examines hybrid volatility methods rather than cross-bank dependence in a state-owned segment
Yaakoub et al. [33]	Four international indices	ARIMA, SVR, LSTM	Compares price forecasts and documents market-dependent model rankings
Present study	Four Indonesian state-owned banks, 2010–2025	RF, GB, lagged cross-bank variables, SRI, volatility timing	Evaluates forecasts, contemporaneous stress, and bank-specific economic outcomes in a defined emerging-market segment

3. Methodology

The dataset contains 3,938 common-date observations of daily adjusted closing prices for Bank Rakyat Indonesia (BBRI), Bank Tabungan Negara (BBTN), Bank Mandiri (BMRI), and Bank Negara Indonesia (BBNI) from January 4, 2010, to December 30, 2025. Constructing rolling variables and lags, aligning the systemic indicators, and shifting the one-step-ahead target leave 3,867 modeling observations. The earliest 80% constitute the estimation sample. The final 774 observations, from October 6, 2022, to December 29, 2025, are reserved as a common test sample for all models. Missing observations are not imputed with future information.

Daily log returns are

$$r_{i,t} = \ln(P_{i,t}) - \ln(P_{i,t-1}), \tag{1}$$

where $P_{i,t}$ is the adjusted closing price. The forecast target is the backward-looking 20-trading-day daily volatility,

$$\sigma_{i,t}^{(20)} = \left[\frac{1}{19} \sum_{j=0}^{19} \left(r_{i,t-j} - \bar{r}_{i,t}^{(20)} \right)^2 \right]^{1/2}, \qquad \bar{r}_{i,t}^{(20)} = \frac{1}{20} \sum_{j=0}^{19} r_{i,t-j}. \tag{2}$$

Individual-bank values near 0.02 are therefore expressed in daily units. For the system-wide component, volatility is annualized before cross-sectional averaging: $AV_t = \sqrt{252} N^{-1} \sum_i \sigma_{i,t}^{(20)}$. This transformation accounts for the larger scale of Average Volatility in Table 3.

The SRI is a descriptive market-based index rather than a substitute for balance-sheet measures such as SRISK or conditional tail measures such as CoVaR. It combines annualized average volatility, the mean of the six rolling 20-day pairwise return correlations, and the downside tail of an equally weighted four-bank portfolio. Formally,

$$AC_t = \frac{2}{N(N-1)} \sum_{i<k} \rho_{ik,t}^{(20)}, \qquad TR_t = -Q_{0.05}\left(\{ r_{t-j}^p \}_{j=0}^{19} \right), \qquad r_t^p = \frac{1}{N} \sum_{i=1}^{N} r_{i,t}, \tag{3}$$

where $N = 4$. Each component is standardized using means and standard deviations estimated from the training period only,

$$Z_{\text{train}}(X_t) = \frac{X_t - \mu_{X,\text{train}}}{s_{X,\text{train}}}. \tag{4}$$

For readability, the standardized composite is placed on an index scale,

$$SRI_t = 20 + 10 \left\{ \frac{Z_{\text{train}}(AV_t) + Z_{\text{train}}(AC_t) + Z_{\text{train}}(TR_t)}{3} \right\}. \tag{5}$$

The base value and scale change the presentation of the index but not its ranking or its use in the forecasting models. Equal weights retain a transparent interpretation. Robustness checks reconstruct the standardized components using a principal-component index estimated on the training sample and an inverse-volatility-weighted index. We also estimate static 5% ΔCoVaR by regressing the equal-weight banking-system return on each bank return using quantile regression. ΔCoVaR is the difference between the fitted system tail return when the bank is at its 5% VaR and when it is at its median state. The regression is estimated by iterative weighted least squares in `statsmodels` 0.14.6. Slope inference uses the software's heteroskedasticity-robust asymptotic covariance estimator with an Epanechnikov kernel and Hall–Sheather bandwidth. SRISK is not estimated because the price dataset lacks the liabilities and market-capitalization inputs required for that measure.

All predictors are observed at forecast origin t, and the target is $\sigma_{i,t+1}^{(20)}$. "Lag 1" therefore refers to the latest available observation, $\sigma_{i,t}^{(20)}$, rather than $\sigma_{i,t-1}^{(20)}$. The predictor set includes volatility lags of 1, 5, 10, and 20 days; current and lagged returns and squared returns; cross-bank volatility; lagged AV, AC, and SRI; a 5-to-20-day volatility ratio; and moving-average crossovers. Hyperparameters are selected from strictly ordered expanding-window folds without random shuffling.

The event-regime analysis relies on a prespecified calendar rather than an SRI-derived threshold. The European sovereign-debt and COVID-19 labels follow an ECB chronology that dates these episodes to 2010–2012 and early 2020 onward [36]. The taper-tantrum window begins with the repricing of U.S. monetary-policy expectations in May 2013, which tightened financial conditions across emerging markets [37]. The 2022–2023 interval covers the global inflation surge and associated monetary tightening documented by the IMF [38]. The corresponding sample windows are January 4, 2010–December 28, 2012; May 1–December 30, 2013; February 3–December 30, 2020; and January 3, 2022–December 29, 2023. Their exact endpoints are operational definitions aligned with the available trading calendar rather than official crisis dates. These intervals contain 1,605 raw observations. After construction of the modeling sample, 1,535 observations are classified as stress periods and 2,332 as normal periods. The classification supports descriptive comparisons and does not assume equal stress intensity throughout each interval.

RF and GB are implemented as CART variance-reduction ensembles in the supplied Python replication code. Candidate configurations are evaluated by expanding-window RMSE, with the selected values reported in Table 2.

3.1. Random Forest Regression

Random Forest estimates decision trees on bootstrap samples and averages their predictions:

$$\hat{y}_{t+1} = \frac{1}{B}\sum_{b=1}^{B} f_b(X_t), \tag{6}$$

where $f_b(\cdot)$ denotes the prediction of tree b, X_t is the vector of lagged predictors, and B is the number of trees.

3.2. Gradient Boosting Regression

Gradient Boosting adds trees sequentially, with each new learner fitted to the residuals of the existing ensemble:

$$F_m(X) = F_{m-1}(X) + \nu \cdot h_m(X), \tag{7}$$

where $h_m(X)$ is the weak learner at iteration m and ν is the learning rate.

The benchmark set comprises persistence, an expanding historical mean, a moving average, HAR/HAR-X, GARCH(1,1), EGARCH(1,1), and GJR-GARCH(1,1). To match the rolling target, each GARCH conditional-variance forecast is combined with the preceding 19 observed returns to obtain the expected next 20-day sample standard deviation. We calculate R^2, RMSE, MAE, and MAPE on the same 774 test dates. For the Diebold–Mariano test, the loss differential is $d_t = e^2_{m,t} - e^2_{RW,t}$, so negative values favor model m. Newey–West HAC standard errors use 19 lags to account for overlap in the 20-day targets.

Table 2. Selected Machine-Learning Hyperparameters

Bank	Model	Trees	Depth	Min. leaf	Learning rate
BBRI	RF	16	6	20	–
BBTN	RF	16	6	20	–
BMRI	RF	16	6	20	–
BBNI	RF	16	6	20	–
BBRI/BMRI	GB	30	3	20	0.05
BBTN/BBNI	GB	24	2	15	0.08

Expanding-window validation selects the same RF configuration for all four banks: 16 trees, a maximum depth of 6, and a minimum leaf size of 20. The preferred GB specification varies across banks. BBRI and BMRI use 30 trees with depth 3 and a learning rate of 0.05, whereas BBTN and BBNI use 24 shallower trees with depth 2 and a learning rate of 0.08. The complete selected configurations are reported in Table 2; their limited depth and tree counts constrain model complexity relative to the available sample.

For the economic evaluation, exposure equals one when the RF forecast falls below the median fitted forecast in the training sample and zero otherwise. The signal formed at t is applied to the return at $t + 1$. The cash position earns a fixed 2% annually, exposure is evaluated daily, and each change in position incurs a cost of 10 basis points. The 2% rate is an ex-ante common benchmark rather than an estimate of the time-varying Indonesian risk-free rate; holding it constant across banks isolates differences generated by the forecast signals and turnover. Economic results are therefore conditional on this cash-return assumption. Reported outcomes include turnover, gross and net Sharpe ratios, cumulative return, and maximum drawdown.

Mean daily rolling volatility ranges from approximately 0.019 for BBRI and BMRI to 0.023 for BBTN, which also has the largest standard deviation (0.0099). Each bank series is right-skewed and heavy-tailed, consistent with occasional high-volatility episodes. The SRI has a mean of 21.92 and a standard deviation of 10.81, while the mean pairwise correlation of 0.6397 indicates substantial comovement within the four-bank sample. The full descriptive statistics are reported in Table 3.

Table 3. Descriptive Statistics of Volatility and Systemic Risk

Variable	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
BBRI Rolling Volatility	0.0192	0.0084	0.0046	0.0710	1.77	6.38
BBTN Rolling Volatility	0.0225	0.0099	0.0049	0.0775	1.56	3.80
BMRI Rolling Volatility	0.0194	0.0082	0.0058	0.0718	1.95	7.09
BBNI Rolling Volatility	0.0195	0.0084	0.0054	0.0697	1.88	5.66
Average Correlation	0.6397	0.1198	0.2966	0.8845	-0.48	-0.14
Average Volatility	0.3286	0.1074	0.1616	0.8748	1.68	3.94
Systemic Risk Index	21.92	10.81	5.55	72.88	1.57	3.38

4. Empirical Results

4.1. Correlation Structure and Interconnectedness

Pairwise correlations summarize volatility comovement within the sample. The largest coefficients are observed for BBRI–BMRI (approximately 0.83), BBRI–BBNI (approximately 0.82), and BMRI–BBNI (approximately 0.82). Correlations involving BBTN are lower, at approximately 0.64–0.66. These estimates indicate comovement but do not identify causal contagion or the direction of shock transmission.

Correlations between individual-bank volatility and the SRI range from approximately 0.66 to 0.74. Average Volatility and Average Correlation correlate with the SRI at approximately 0.97 and 0.82, respectively. Because both variables enter the composite, these coefficients indicate internal consistency rather than external construct validity or contagion.

4.2. Event-Defined Stress versus Normal Periods

After construction of the rolling variables and lags, the sample contains 1,535 stress observations and 2,332 normal observations. Mean volatility is higher during the prespecified stress intervals for all four banks (Table 4). Because the dates are defined independently of the SRI, this comparison does not mechanically select observations with high index values.

Table 4. Event-Defined Stress-Regime Analysis

Variable	Normal Period	Stress Period
Observations	2,332	1,535
BBRI Volatility	0.018043	0.020879
BBTN Volatility	0.021924	0.023457
BMRI Volatility	0.017922	0.021606
BBNI Volatility	0.018608	0.020749
Average Correlation	0.641832	0.636475
Average Volatility	0.311302	0.354756
Systemic Risk Index	20.526232	24.030824

Average Volatility increases from 0.311302 in normal periods to 0.354756 in stress periods, whereas Average Correlation declines slightly from 0.641832 to 0.636475. In these event windows, higher system-wide volatility is therefore not accompanied by a uniform increase in contemporaneous correlation.

4.3. Systemic Risk Dynamics

The SRI and annualized average volatility reach their largest joint peak during the 2020 pandemic interval, as shown by the daily series in Figure 1. Average Correlation remains persistent but does not increase in every event window, consistent with the regime comparison in Table 4.

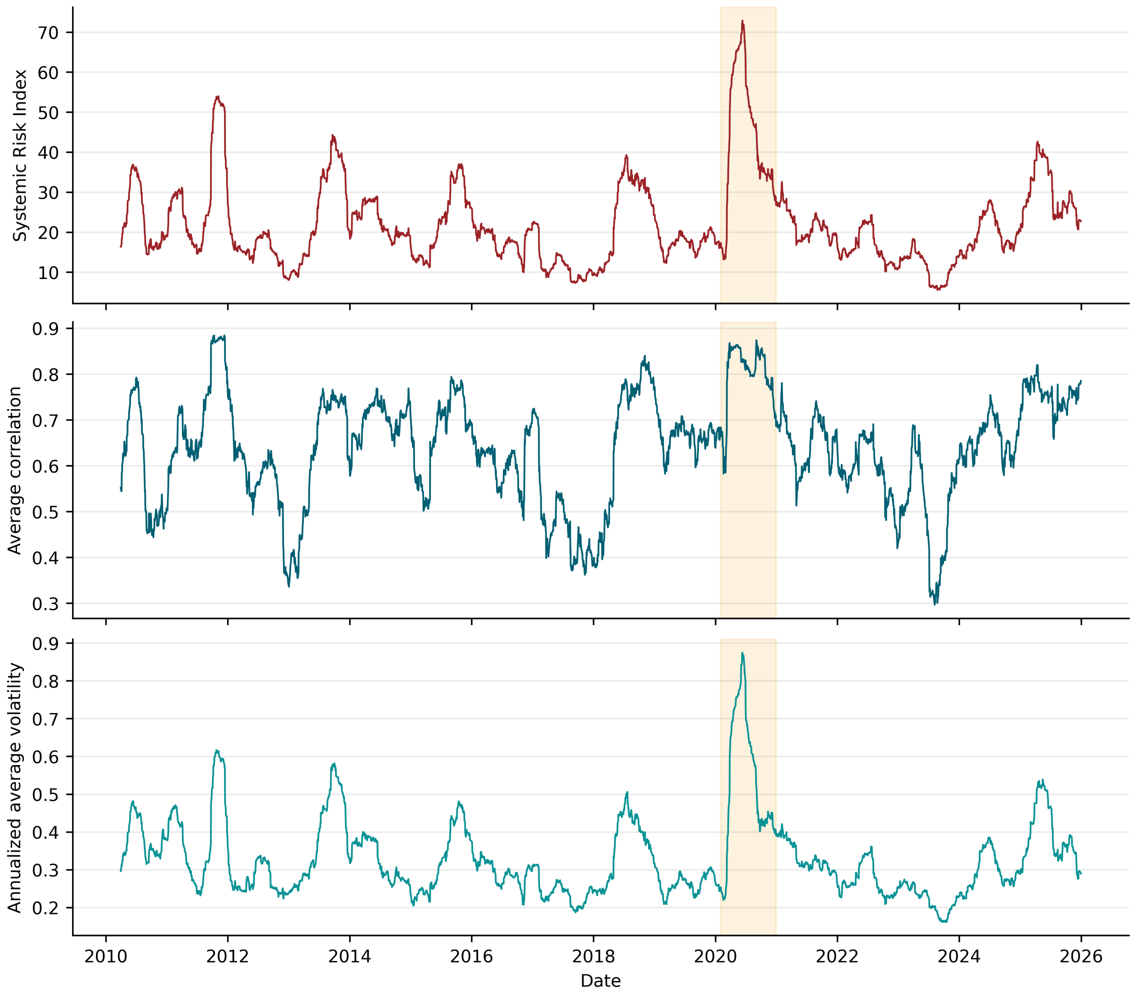

Figure 1. Systemic risk, average correlation, and average volatility. The shaded interval denotes the 2020 pandemic event window.

4.4. Unified Forecast Comparison

On the common 774-day test sample, the Random Walk outperforms RF and GB for every bank, while the GARCH-family models improve on persistence. The best GARCH-family specification yields R^2 values of 0.9793–0.9838 and MAPE values of 3.18–4.16% across banks. The corresponding RF ranges are 0.8929–0.9459 for R^2 and 6.58–15.93% for MAPE. Table 5 reports the complete model-by-bank comparison.

All forecasts are evaluated in the same units because each conditional GARCH variance forecast is converted to the expected next 20-day rolling standard deviation using the preceding 19 observed returns. The resulting R^2 ranking, displayed in Figure 2, places the GARCH-family models above persistence and both tree-based models.

BBTN remains the least accurately forecast bank under the ML models, with a GB R^2 of 0.8631 and an RF MAPE of 15.93%. Its return volatility during 2020–2025 is 0.0256, and its maximum 20-day volatility is 0.0739. These estimates document greater forecasting difficulty but do not identify its economic source.

Table 5. Unified Out-of-Sample Forecast Performance

Bank	Model	R^2	RMSE	MAE	MAPE (%)
BBRI	RF	0.9258	0.001701	0.001118	6.58
	GB	0.9169	0.001801	0.001274	8.31
	Random Walk	0.9661	0.001151	0.000622	3.78
	GARCH	0.9804	0.000874	0.000531	3.17
	EGARCH	0.9805	0.000873	0.000548	3.33
	GJR-GARCH	0.9806	0.000870	0.000532	3.18
BBTN	RF	0.8929	0.002755	0.002006	15.93
	GB	0.8631	0.003116	0.002183	18.66
	Random Walk	0.9650	0.001576	0.000826	4.47
	GARCH	0.9804	0.001179	0.000709	4.25
	EGARCH	0.9807	0.001171	0.000709	4.16
	GJR-GARCH	0.9805	0.001177	0.000705	4.27
BMRI	RF	0.9256	0.001733	0.001192	6.76
	GB	0.8941	0.002067	0.001529	9.21
	Random Walk	0.9598	0.001274	0.000690	4.02
	GARCH	0.9793	0.000914	0.000560	3.20
	EGARCH	0.9794	0.000913	0.000566	3.26
	GJR-GARCH	0.9793	0.000914	0.000559	3.19
BBNI	RF	0.9459	0.001609	0.001196	7.84
	GB	0.9120	0.002054	0.001580	11.48
	Random Walk	0.9698	0.001203	0.000691	4.25
	GARCH	0.9836	0.000887	0.000560	3.42
	EGARCH	0.9838	0.000882	0.000560	3.42
	GJR-GARCH	0.9837	0.000884	0.000559	3.43

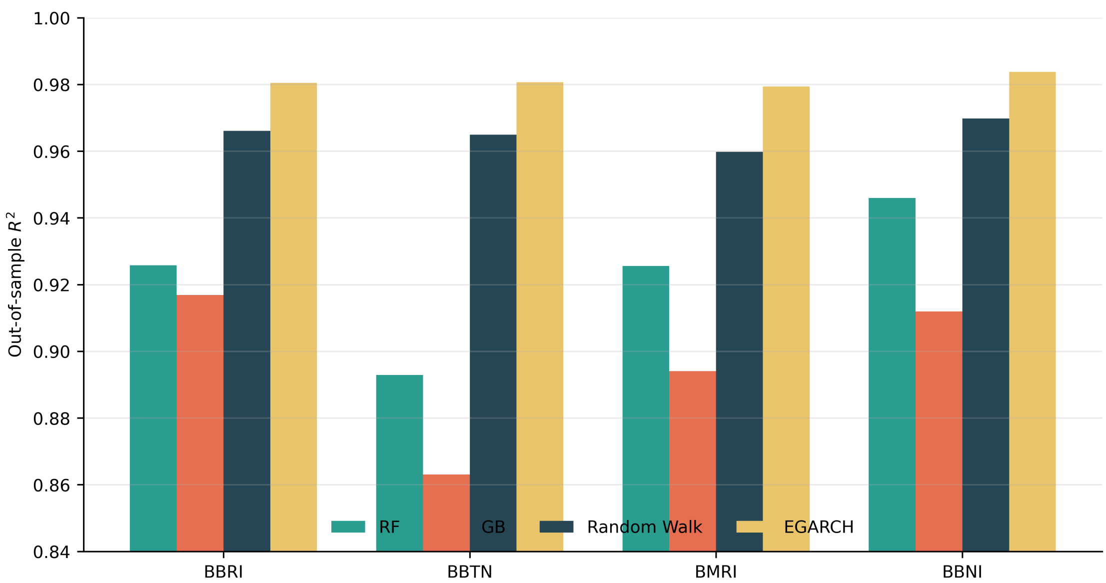

Figure 2. Out-of-sample R^2 comparison on the common test period.

4.5. *Feature Importance and Systemic Role*

Each bank's own one-day volatility is the most influential predictor, particularly in GB, and removing it substantially lowers R^2. Excluding the systemic variables produces smaller and mixed changes. Removing BBRI Lag 1 from the other-bank RF models changes R^2 from 0.8929 to 0.8891 for BBTN, from 0.9256 to 0.9318 for

BMRI, and from 0.9459 to 0.9384 for BBNI. The importance and ablation estimates in Table 6 therefore indicate some cross-bank information in BBRI but do not support a uniform bellwether interpretation.

Table 6. Feature Importance and Ablation Results

Bank	Model	Own-Lag-1 Imp.	Full R^2	No systemic R^2	No own Lag 1 R^2
BBRI	RF	0.3336	0.9258	0.9372	0.8714
	GB	0.9475	0.9169	0.9169	0.8056
BBTN	RF	0.3878	0.8929	0.8976	0.8169
	GB	0.9944	0.8631	0.8615	0.7198
BMRI	RF	0.3220	0.9256	0.9346	0.8679
	GB	0.9587	0.8941	0.8941	0.7774
BBNI	RF	0.3940	0.9459	0.9389	0.8876
	GB	0.9442	0.9120	0.9119	0.7848

The systemic indicators have heterogeneous incremental forecasting content. Removing them improves RF accuracy for three banks but reduces it for BBNI, while GB performance is almost unchanged. The held-out ablation evidence is more informative than impurity rankings alone and shows that high contemporaneous correlations with the SRI need not translate into one-day-ahead predictive gains.

4.6. *SRI and Tail-Risk Robustness*

The first principal component explains 76.73% of variation in the three SRI components and correlates 0.999 with the equal-weight composite. The inverse-volatility index correlates 0.947 with the equal-weight index, and the reconstructed equal-weight index correlates 0.855 with the supplied SRI series. The ranking of stress periods is therefore not specific to equal weighting. Table 7 reports ΔCoVaR as a separate measure of tail dependence.

Table 7. ΔCoVaR Robustness Results

Bank	Bank VaR (5%)	CoVaR distress	ΔCoVaR	Quantile slope
BBRI	-0.0316	-0.0385	-0.0234	0.7396
BBTN	-0.0359	-0.0378	-0.0207	0.5786
BMRI	-0.0325	-0.0404	-0.0248	0.7640
BBNI	-0.0326	-0.0397	-0.0240	0.7356

All quantile slopes are positive and statistically significant at $p < 0.0001$. BMRI has the largest absolute ΔCoVaR (-0.0248), followed by BBNI and BBRI; BBTN has the smallest value (-0.0207). These estimates indicate downside dependence but do not imply that the SRI is equivalent to CoVaR or SRISK.

4.7. *Window and Subperiod Robustness*

Forecast sensitivity differs markedly across the 5-, 10-, 20-, and 60-day volatility windows. The largest improvement over persistence occurs at 20 days, where the GARCH family reduces average RMSE by 26.29%. Improvements at the other windows are smaller and are generally produced by HAR or HAR-X. The numerical comparison and its graphical representation appear in Table 8 and Figure 3, respectively.

This window sensitivity establishes that the model ranking is horizon dependent. Forecast stability also varies over time: RF R^2 ranges from 0.8587 to 0.9079 across banks and subperiods. In 2020–2025, GB R^2 falls to 0.8392 for BBRI, 0.8709 for BBTN, and 0.8527 for BMRI. By contrast, BBNI RF R^2 increases from 0.8727 in 2013–2014 to 0.9079 in 2020–2025.

Table 8. Average RMSE across Alternative Volatility Windows

Window	Random Walk	Best alternative	Reduction (%)	Typical best model
5	0.005302	0.004989	5.91	HAR-X
10	0.002616	0.002558	2.21	HAR-X
20	0.001301	0.000959	26.29	GARCH family
60	0.000454	0.000446	1.74	HAR/HAR-X

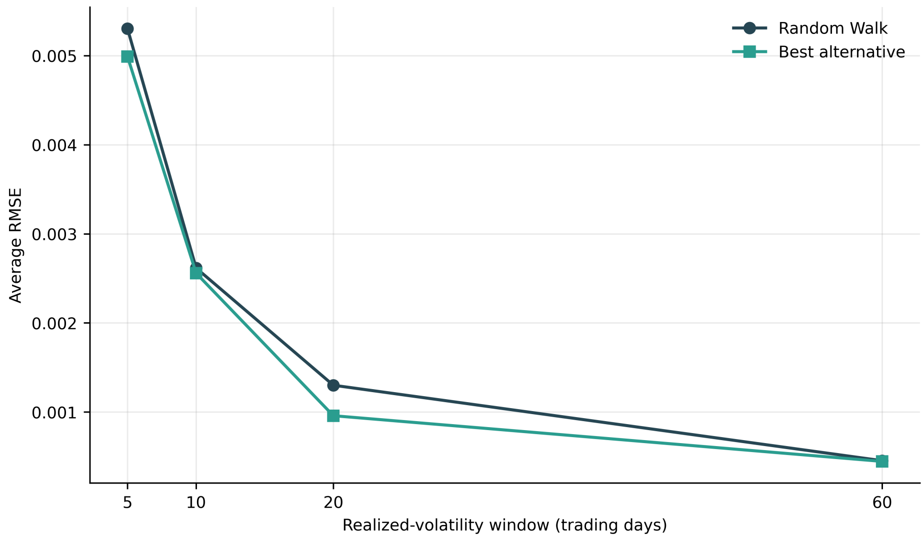

Figure 3. Forecast sensitivity to the rolling-volatility window.

4.8. Statistical Significance

The HAC-corrected DM statistics are positive for RF and GB, indicating lower loss under the Random Walk, and negative for EGARCH, indicating lower loss under EGARCH. These significance tests, reported in Table 9, agree with the point-forecast rankings.

Table 9. HAC-Corrected Diebold–Mariano Tests against Random Walk

Bank	Model	HAC statistic	p-value	Favored model
BBRI	RF	2.766	0.0057	Random Walk
	GB	2.767	0.0057	Random Walk
	EGARCH	-2.856	0.0043	EGARCH
BBTN	RF	3.214	0.0013	Random Walk
	GB	3.084	0.0020	Random Walk
	EGARCH	-5.790	< 0.0001	EGARCH
BMRI	RF	4.811	< 0.0001	Random Walk
	GB	2.794	0.0052	Random Walk
	EGARCH	-4.204	< 0.0001	EGARCH
BBNI	RF	6.546	< 0.0001	Random Walk
	GB	4.567	< 0.0001	Random Walk
	EGARCH	-5.669	< 0.0001	EGARCH

Persistence is the strongest nonparametric benchmark and outperforms both tree models. It is not the best overall forecast, however, once the GARCH-family forecasts are expressed in the same 20-day volatility units.

4.9. Economic Significance

The ex-ante RF timing rule fixes its exposure threshold using fitted forecasts from the training sample. Cash earns the fixed 2% annual benchmark specified above, and net returns deduct 10 basis points for each change in exposure. Table 10 reports the resulting Sharpe ratios, active exposure, and annual turnover.

Table 10. Ex-Ante Economic Significance with Transaction Costs

Bank	Buy-and-hold Sharpe	Gross Sharpe	Net Sharpe	Active ratio	Annual turnover
BBRI	-0.035	0.341	0.295	0.625	8.79
BBTN	-0.019	-0.443	-0.505	0.561	13.02
BMRI	0.241	0.542	0.488	0.623	11.40
BBNI	0.104	0.185	0.140	0.673	8.79

Relative to buy-and-hold, net Sharpe ratios are higher for BBRI, BMRI, and BBNI but lower for BBTN. BBTN also has the highest annual turnover (13.02), indicating that weaker forecasts coincide with greater trading costs. These results are conditional on the specified backtest and do not constitute investment advice.

5. Discussion

The comparison qualifies claims of general machine-learning superiority in volatility forecasting. Although ML models have performed well in larger developed-market samples [20], nonlinear flexibility does not offset strong persistence in this four-bank setting. GARCH-family forecasts perform best after being mapped to the rolling 20-day target, which is consistent with evidence that rankings vary across markets, targets, and hybrid specifications [16, 34]. The result also accords with SOIC evidence that model performance depends on the persistence properties of the underlying series [32]. Because adjacent 20-day targets share 19 observations, their high R^2 values partly reflect mechanical persistence. The HAC correction addresses serial dependence in forecast-loss inference, but the magnitude of R^2 should still be interpreted in light of this overlap.

Ablation identifies own-volatility persistence as the main source of ML accuracy: removing own Lag 1 reduces performance for every bank, consistent with conventional and machine-learning volatility designs [9, 20]. Systemic variables and BBRI Lag 1 contribute less consistently. This finding separates connectedness from forecast content. Network measures describe dependence and potential transmission channels [10, 11], but they do not imply that one bank must improve every other bank's short-horizon forecast. BBRI's importance in selected models is therefore evidence of conditional predictive association, not uniform bellwether status or causal transmission [30].

The SRI results clarify its empirical role. Similar rankings under equal weighting, PCA, and inverse-volatility weighting indicate that the main stress episodes are not produced by a single weighting choice. Positive quantile slopes and negative ΔCoVaR estimates are consistent with common market movements and institutional connectedness contributing to aggregate financial risk [5, 6]. Their limited contribution to one-day-ahead forecasts is not contradictory: a variable can measure the contemporaneous state of the system without leading future bank volatility. Evidence from concentrated banking systems similarly shows that dependence need not imply a stable predictive direction [24].

Economic value is also bank dependent. After switching costs, the timing rule improves Sharpe ratios for BBRI, BMRI, and BBNI but not for BBTN. This heterogeneity accords with evidence that volatility forecasts create portfolio value only through the joint effect of forecast quality, exposure rules, and trading costs [35]. BBTN combines the weakest ML accuracy with the highest turnover, so its negative outcome is consistent with the mechanics of the strategy rather than evidence against volatility timing in every bank.

Several design choices delimit these conclusions. The sample does not represent private, foreign, Islamic, or regional Indonesian banks. The static ΔCoVaR estimates capture unconditional sample-level tail dependence rather than time-varying systemic contributions, and the absence of liabilities and market-capitalization data precludes SRISK estimation. A compact hyperparameter grid limits model-search overfitting but does not exhaust the set of possible architectures. Daily observations cannot identify intraday transmission, while the broad event windows support descriptive regime comparisons rather than precise crisis dating.

6. Conclusion

For the four Indonesian state-owned banks examined from 2010 to 2025, RF and GB capture nonlinear and cross-bank associations but do not outperform persistence. GARCH-family models produce the most accurate 20-day volatility forecasts and significantly reduce loss relative to the Random Walk. Systemic indicators describe common stress and downside dependence, yet add limited and heterogeneous short-horizon forecast information. Their principal value in this setting is therefore contemporaneous monitoring rather than uniform early warning.

The economic results reinforce the institutional boundary of the analysis: volatility timing improves net Sharpe ratios for three banks but fails for BBTN. Broader samples, balance-sheet inputs for SRISK, release-date-aligned macroeconomic variables, and intraday data are needed to assess whether these patterns extend beyond the listed state-owned segment. Larger samples would also permit deeper architectures without making inference excessively dependent on model selection.

Data and Code Availability

The underlying daily adjusted closing-price data for BBRI.JK, BBTN.JK, BMRI.JK, and BBNI.JK are publicly available from Yahoo Finance for the period January 2010–December 2025. The Python analysis code, model settings, and supporting files required to reproduce the reported results are available from the corresponding author upon reasonable request.

Funding

This research was supported by Universitas Singaperbangsa Karawang.

Author Contributions

Nono Heryana: Conceptualization, methodology, software, formal analysis, data curation, visualization, and writing–original draft. Nugraha Nugraha: Validation, investigation, and writing–review and editing. Maya Sari: Conceptualization, methodology, validation, resources, supervision, project administration, funding acquisition, writing–original draft, and writing–review and editing. Disman Disman: Methodology, validation, supervision, and writing–review and editing. Toni Heryana: Software, formal analysis, visualization, and writing–review and editing. Rini Mayasari: Validation, investigation, resources, data curation, and writing–review and editing. All the authors have read and approved the final version of the manuscript.

Conflict of Interest

The authors declare no conflicts of interest.

REFERENCES

1. M. Naili and Y. Lahrichi, "Banks' credit risk, systematic determinants and specific factors: recent evidence from emerging markets," *Heliyon*, vol. 8, no. 2, p. e08960, Feb. 2022, doi: 10.1016/j.heliyon.2022.e08960.
2. S. Ullah, A. Ullah, and M. Zaman, "Nexus of governance, macroeconomic conditions, and financial stability of banks: a comparison of developed and emerging countries," *Financ Innov*, vol. 10, no. 1, p. 30, Jan. 2024, doi: 10.1186/s40854-023-00542-x.
3. M. R. Rabbani, U. Kayani, H. S. Bawazir, and I. T. Hawaldar, "A commentary on emerging markets banking sector spillovers: Covid-19 vs GFC pattern analysis," *Heliyon*, vol. 8, no. 3, p. e09074, Mar. 2022, doi: 10.1016/j.heliyon.2022.e09074.
4. G. Taneva-Angelova and D. Granchev, "Deep Learning and Transformer Architectures for Volatility Forecasting: Evidence from U.S. Equity Indices," *JRFM*, vol. 18, no. 12, p. 685, Dec. 2025, doi: 10.3390/jrfm18120685.
5. M. Billio, M. Getmansky, A. W. Lo, and L. Pelizzon, "Econometric measures of connectedness and systemic risk in the finance and insurance sectors," *Journal of Financial Economics*, vol. 104, no. 3, pp. 535–559, Jun. 2012, doi: 10.1016/j.jfineco.2011.12.010.
6. S. Giglio, B. Kelly, and S. Pruitt, "Systemic risk and the macroeconomy: An empirical evaluation," *Journal of Financial Economics*, vol. 119, no. 3, pp. 457–471, Mar. 2016, doi: 10.1016/j.jfineco.2016.01.010.
7. S. Basak, S. Kar, S. Saha, L. Khaidem, and S. R. Dey, "Predicting the direction of stock market prices using tree-based classifiers," *The North American Journal of Economics and Finance*, vol. 47, pp. 552–567, Jan. 2019, doi: 10.1016/j.najef.2018.06.013.
8. E. K. Ampomah, Z. Qin, and G. Nyame, "Evaluation of Tree-Based Ensemble Machine Learning Models in Predicting Stock Price Direction of Movement," *Information*, vol. 11, no. 6, p. 332, Jun. 2020, doi: 10.3390/info11060332.
9. Z. Qiu, C. Kownatzki, F. Scalzo, and E. S. Cha, "Historical Perspectives in Volatility Forecasting Methods with Machine Learning," *Risks*, vol. 13, no. 5, p. 98, May 2025, doi: 10.3390/risks13050098.
10. F. X. Diebold and K. Yılmaz, "On the network topology of variance decompositions: Measuring the connectedness of financial firms," *Journal of Econometrics*, vol. 182, no. 1, pp. 119–134, Sep. 2014, doi: 10.1016/j.jeconom.2014.04.012.
11. M. Demirer, F. X. Diebold, L. Liu, and K. Yilmaz, "Estimating global bank network connectedness," *J of Applied Econometrics*, vol. 33, no. 1, pp. 1–15, Jan. 2018, doi: 10.1002/jae.2585.
12. H. A. Mieg, "Volatility as a Transmitter of Systemic Risk: Is there a Structural Risk in Finance?," *Risk Analysis*, vol. 42, no. 9, pp. 1952–1964, Sep. 2022, doi: 10.1111/risa.13564.
13. S. Song and H. Li, "Early Warning of Systemic Risk Via Dynamic Co-Jump Network Topology: Evidence from Emerging and Advanced Financial Markets," *Emerging Markets Finance and Trade*, vol. 62, no. 7, pp. 2439–2455, May 2026, doi: 10.1080/1540496X.2025.2573437.
14. C. Tudor, A. Girlovan, G. R. Saiu, and D. D. Guse, "Asymmetric Shocks and Pension Fund Volatility: A GARCH Approach with Macroeconomic Predictors to an Unexplored Emerging Market," *Mathematics*, vol. 13, no. 7, p. 1134, Mar. 2025, doi: 10.3390/math13071134.
15. J. Caiado and F. Lúcio, "Stock market forecasting accuracy of asymmetric GARCH models during the COVID-19 pandemic," *The North American Journal of Economics and Finance*, vol. 68, p. 101971, Sep. 2023, doi: 10.1016/j.najef.2023.101971.
16. V. Chung, J. Espinoza, and R. Quispe, "Forecasting Financial Volatility Under Structural Breaks: A Comparative Study of GARCH Models and Deep Learning Techniques," *JRFM*, vol. 18, no. 9, p. 494, Sep. 2025, doi: 10.3390/jrfm18090494.
17. S. Ben Salem and O. El Aoun, "Banking index volatility and spillover effects in G7 and BRICS economies: a quantile connectedness approach," *SEF*, vol. 42, no. 4, pp. 836–864, Jul. 2025, doi: 10.1108/SEF-01-2025-0046.
18. S. Gu, B. Kelly, and D. Xiu, "Empirical Asset Pricing via Machine Learning," *The Review of Financial Studies*, vol. 33, no. 5, pp. 2223–2273, May 2020, doi: 10.1093/rfs/hhaa009.
19. A. Weigand, "Machine learning in empirical asset pricing," *Financ Mark Portf Manag*, vol. 33, no. 1, pp. 93–104, Mar. 2019, doi: 10.1007/s11408-019-00326-3.
20. K. Christensen, M. Siggaard, and B. Veliyev, "A Machine Learning Approach to Volatility Forecasting," *Journal of Financial Econometrics*, vol. 21, no. 5, pp. 1680–1727, Nov. 2023, doi: 10.1093/jjfinec/nbac020.
21. J. El Gallab, N. Abdelbary, M. Elfar, F. Abouelghit, B. Rady, and R. Bairakdar, "Towards a Machine Learning-Based Approach to Predicting Stock Price Volatility and Its Associated Risk in Egypt," in *Machine Learning and Soft Computing*, vol. 2487, L. Huang, Ed., in *Communications in Computer and Information Science*, vol. 2487. Singapore: Springer Nature Singapore, 2025, pp. 157–170, doi: 10.1007/978-981-96-6400-9_12.
22. T. Sharma, S. K. Prasad, S. Prasad, I. Verma, and A. Sharma, "Forecasting Stock Market Volatility Using XGBoost: A Time Series Analysis," in *2024 International Conference on Communication, Control, and Intelligent Systems (CCIS)*, Mathura, India: IEEE, Dec. 2024, pp. 1–6, doi: 10.1109/CCIS63231.2024.10932089.
23. T. Adrian, N. Boyarchenko, and D. Giannone, "Vulnerable Growth," *American Economic Review*, vol. 109, no. 4, pp. 1263–1289, Apr. 2019, doi: 10.1257/aer.20161923.
24. Q. Shi, X. Sun, and Y. Jiang, "Concentrated commonalities and systemic risk in China's banking system: A contagion network approach," *International Review of Financial Analysis*, vol. 83, p. 102253, Oct. 2022, doi: 10.1016/j.irfa.2022.102253.
25. T. Matsuoka, "Banks and liquidity crises in emerging market economies," *Journal of Economic Dynamics and Control*, vol. 94, pp. 43–62, Sep. 2018, doi: 10.1016/j.jedc.2018.07.002.
26. T. Wang, S. Zhao, G. Zhu, and H. Zheng, "A machine learning-based early warning system for systemic banking crises," *Applied Economics*, vol. 53, no. 26, pp. 2974–2992, Jun. 2021, doi: 10.1080/00036846.2020.1870657.
27. W. E. AboElnasr, M. A. Zahran, and M. M. Abdelsalam, "Risk warning system for financial crises using multifractal analysis and dictionary learning," *Chaos, Solitons & Fractals*, vol. 201, p. 117194, Dec. 2025, doi: 10.1016/j.chaos.2025.117194.
28. S. K. Sahu, A. Mokhade, and N. D. Bokde, "An Overview of Machine Learning, Deep Learning, and Reinforcement Learning-Based Techniques in Quantitative Finance: Recent Progress and Challenges," *Applied Sciences*, vol. 13, no. 3, p. 1956, Feb. 2023, doi: 10.3390/app13031956.
29. C. R. Harvey, Y. Liu, and H. Zhu, "… and the Cross-Section of Expected Returns," *Rev. Financ. Stud.*, vol. 29, no. 1, pp. 5–68, Jan. 2016, doi: 10.1093/rfs/hhv059.

30. G. Xu and W. Gao, “Financial Risk Contagion in Stock Markets: Causality and Measurement Aspects,” *Sustainability*, vol. 11, no. 5, p. 1402, Mar. 2019, doi: 10.3390/su11051402.
31. H. Oukhouya, H. Kadiri, K. El Himdi, and R. Guerbaz, “Forecasting International Stock Market Trends: XGBoost, LSTM, LSTM-XGBoost, and Backtesting XGBoost Models,” *Statistics, Optimization & Information Computing*, vol. 12, no. 1, pp. 200–209, 2024, doi: 10.19139/soic-2310-5070-1822.
32. Y. Bakkali, M. El Merzguioui, and A. Akharif, “Enhancing Volatility Prediction: Comparison Study Between Persistent and Anti-persistent Financial Series,” *Statistics, Optimization & Information Computing*, vol. 12, no. 4, pp. 1042–1060, 2024, doi: 10.19139/soic-2310-5070-2021.
33. A. Yaakoub, H. Oukhouya, M. Elhia, T. Zari, and R. Guerbaz, “Deep Learning for Financial Time Series: Does LSTM Outperform ARIMA and SVR in International Stock Market Predictions?” *Statistics, Optimization & Information Computing*, vol. 14, no. 5, pp. 2181–2199, 2025, doi: 10.19139/soic-2310-5070-2602.
34. N. Oualla, M. S. Chiadmi, and Y. Lamrani Alaoui, “Advancing Volatility Forecasting in Financial Indices: Integrating GARCH Models, Multifractal Indicators, and Deep Learning,” *Statistics, Optimization & Information Computing*, vol. 15, no. 4, pp. 2696–2714, 2026, doi: 10.19139/soic-2310-5070-3018.
35. D. Chun, H. Cho, and D. Ryu, “Volatility forecasting and volatility-timing strategies: A machine learning approach,” *Research in International Business and Finance*, vol. 75, p. 102723, 2025, doi: 10.1016/j.ribaf.2024.102723.
36. K. Hobelsberger, C. Kok, and F. P. Mongelli, “A tale of three crises: synergies between ECB tasks,” *ECB Occasional Paper Series*, no. 305, Sep. 2022, revised Jun. 2023, doi: 10.2866/622341.
37. N. Harikrishnan, B. Silk, and E. Yoldas, “U.S. Interest Rates and Emerging Market Currencies: Taking Stock 10 Years After the Taper Tantrum,” *FEDS Notes*, Board of Governors of the Federal Reserve System, Oct. 4, 2023, doi: 10.17016/2380-7172.3385.
38. International Monetary Fund, *World Economic Outlook: Navigating Global Divergences*. Washington, DC: International Monetary Fund, Oct. 2023, doi: 10.5089/9798400235801.081.